



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : H04Q 11/04, H04L 12/56		A1	(11) International Publication Number: WO 97/43869 (43) International Publication Date: 20 November 1997 (20.11.97)
<p>(21) International Application Number: PCT/US97/07839</p> <p>(22) International Filing Date: 9 May 1997 (09.05.97)</p> <p>(30) Priority Data: 08/648,556 15 May 1996 (15.05.96) US</p> <p>(71) Applicant: CISCO SYSTEMS, INC. [US/US]; 170 West Tasman Drive, San Jose, CA 95134 (US).</p> <p>(72) Inventors: HUGHES, David, A.; 900 High School Way #2230, Mountain View, CA 94041 (US). KLAUSMEIER, Daniel, E.; 1130 Pimenta Avenue, Sunnyvale, CA 94087 (US).</p> <p>(74) Agents: VINCENT, Lester, J. et al.; Blakely, Sokoloff, Taylor & Zafman, 7th floor, 12400 Wilshire Boulevard, Los Angeles, CA 90025-1026 (US).</p>		<p>(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: METHOD AND APPARATUS FOR PER TRAFFIC FLOW BUFFER MANAGEMENT</p> <p>(57) Abstract</p> <p>A method of managing oversubscription of a common buffer resource shared by a number of traffic flows in a cell switching network in response to the utilization of the common buffer resource. A buffer utilization threshold is established for each of the traffic flows. As new cells arrive, the global usage of the buffer resource is monitored. As the buffer utilization increases, the thresholds for each of the traffic flows are dynamically adjusted based upon the global usage of the buffer. Aggressive buffer allocations are scaled back when necessary, thereby leaving space for traffic flows which are relatively empty. In one embodiment, the thresholds are coded in mantissa and exponent form so that the scaling is accomplished by adjusting the exponent value.</p>			

```

graph TD
    14((New Cell Arrives)) --> 16[Determine Traffic Flow for New Cell]
    16 --> 18[Get Default Thresh. and Flow Cell Count]
    18 --> 20[Get Global Buffer Count]
    20 --> 22[Determine Scaling Factor as a Function of Global Buffer Count]
    22 --> 24[Calculate Dynamic Thresh = Thresh * Scaling Factor]
    24 --> 26{Flow Cell Count > Dynamic Thresh?}
    26 -- N --> 30[Admit Cell and Increment Flow Cell Count and Global Buffer Count]
    26 -- Y --> 28[Drop Cell]
  
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NB	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD AND APPARATUS FOR PER TRAFFIC FLOW BUFFER MANAGEMENT

FIELD OF THE INVENTION

The present invention relates generally to the field of cell switching network communications and, more specifically, to the efficient management of shared buffer resources within such a network.

BACKGROUND

The desire to integrate data, voice, image and video over high speed digital trunks has led to the development of a packet switching technique called cell relay or asynchronous transfer mode (ATM). ATM traffic is switched and multiplexed in fixed length cells and an ATM network typically provides a number of interconnected nodes which are capable of receiving data from other network nodes and forwarding that data through to other network nodes to its ultimate destination. Nodes are interconnected by transmission paths, each of which supports one or more virtual paths. Each virtual path contains one or more virtual channels. Switching can be performed at the transmission path, virtual path or virtual channel level.

Network nodes generally employ buffering schemes to prevent contention for switch resources (e.g., ports). In the past, this has included relatively unsophisticated solutions, such as a first-in-first-out (FIFO) queue at each port. This solution quickly leads to cells being dropped indiscriminately when the volume of network traffic is large. Other schemes involve "per connection" buffering where each logical connection (i.e., virtual path, virtual channel) is allocated its own cell memory. When the number of supported connections is large, however, the sum of the maximum buffer requirements for individual connections may drastically exceed the physical available memory.

If one large buffer resource is to be shared among a number of connections then, some form of buffer management must be employed. In the past, one solution has been to divide the buffer into a number of queues of fixed length and "hard allocate" capacity for each connection. The problem with this solution is that the fixed length queues offer no flexibility depending upon network traffic conditions. In addition, because of size and cost constraints, each queue would have to remain relatively small as a single switch may support thousands of logical

connections. Those network connections with significant amounts of traffic would likely soon fill up their allotted queue and cell dropping would soon result. Another solution has been to oversubscribe the single memory resource and allow each connection to buffer up to a fixed maximum, but where the sum of all the connection maxima exceeds the memory capacity. This alternative relies on the fact that all connections are unlikely to require their maximum buffer space at the same time. Although this condition is true most of the time, it is inevitable that contention for buffer space will result at some point. Once contention does result, cells are dropped indiscriminately, i.e., without regard for whether a connection is already using a significant amount of buffer space or not. A third solution has been to reserve a minimum buffer allocation for each connection with the unallocated space available on a first-come-first-served basis. This allows each connection a guaranteed minimum buffer space. The problem with this solution is that where the number of logical connections runs into the thousands, a very large (i.e., expensive) common buffer is required for any reasonable minimum.

None of the buffer management schemes of the prior art have satisfactorily addressed the problem of per connection buffer management for large numbers of connections. Hence, it would be desirable to have a mechanism for effectively managing the oversubscription of a shared buffer resource.

SUMMARY AND OBJECTS OF THE INVENTION

It is therefore an object of the present invention to provide an improved method for managing the oversubscription of a common communications resource shared by a large number of traffic flows, such as ATM connections.

It is a further object of the present invention to provide an efficient method of buffer management at the connection level of a cell switching data communication network so as to minimize the occurrence of resource overflow conditions.

This and other objects of the invention are achieved by an effective method for managing oversubscription by dynamically changing the maximum buffer space allowed for a particular traffic flow or connection in response to the global utilization of a single buffer resource. A buffer utilization threshold for each of a number of various traffic flows is established. As new cells arrive, the global usage of the buffer resource is monitored. As the buffer fills, the individual

thresholds for the various traffic flows are dynamically scaled based upon the global usage of the buffer. This method allows guaranteed buffer space for sensitive traffic flows despite the oversubscription. Aggressive buffer allocations are scaled back when necessary, thereby leaving space for traffic flows which are using only a small portion of their allocated buffer space. The present invention in effect maintains isolation between well behaved traffic flows, insuring that only flows which are using a disproportionate amount of memory are blocked from storing further cells or packets in the memory when the global resource usage approaches capacity.

In one embodiment, the thresholds are coded in mantissa and exponent form so that the scaling is accomplished by adjusting the exponent. This approach allows a minimum of memory to be used to store the flow thresholds and simplifies the mechanism for scaling the thresholds.

Other objects, features and advantages of the present invention will be apparent from the accompanying drawings and from the detailed description which follows.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements, and in which:

Figure 1a is a flow diagram illustrating the initialization of various parameters according to one embodiment;

Figure 1b is a flow diagram illustrating dynamic per traffic flow buffer management according to one embodiment;

Figure 1c is a flow diagram illustrating a cell service routine according to one embodiment;

Figure 2 is a plot which graphically illustrates the number of cells stored in a common buffer by each of a number of traffic flows;

Figure 3 is a plot similar to the plot shown in Figure 2 that further shows a number of buffer thresholds corresponding to the various traffic flows sharing the common buffer;

Figure 4 is a plot similar to the plot shown in Figure 3 and shows the common buffer utilization at a later time;

-4-

Figure 5 is a plot similar to the plot shown in Figure 4 and illustrates the effect of dynamic buffer threshold scaling for one traffic flow according to one embodiment; and

Figure 6 is a plot similar to the plot shown in Figure 4 and illustrates the effects of dynamic buffer threshold scaling for a different traffic flow according to one embodiment.

DETAILED DESCRIPTION

An improved method and apparatus to efficiently manage a common communications buffer resource shared by a large number of traffic flows, such as a cell memory shared by ATM virtual channels or paths, is described. According to one embodiment, oversubscription of a shared buffer resource is managed by dynamically changing the maximum buffer space allowed for each traffic flow in response to the global utilization of a single shared buffer resource.

Upon review of this specification, those skilled in the art will appreciate that the methods and apparatus to be described are applicable at a number of levels. For example, the methods can be employed at the "per logical connection" level or at the "per quality of service" level, among others. To account for the numerous levels at which the present invention is applicable, the term "traffic flow" is used throughout this specification. Those skilled in the art will appreciate that this term describes the general nature of the levels at which the present invention is applicable. A presently preferred embodiment utilizes the invention at the per logical connection level in managing common buffer resources in ATM network nodes. In this particular case, a traffic flow is associated with the transport of cells on a single logical connection. The particular nature of this embodiment should not, however, be seen as limiting the more general nature and scope of the present invention as set forth in the appended claims.

In a fully integrated voice and data telecommunications network, a variety of switching nodes will be present. Each node is interconnected to other network nodes by a variety of transmission paths. The apparent capacity of these transmission paths is increased using virtual connections. In other words, rather than committing specific resources to a given source-destination pair, each node connects a source-destination pair only when information, in the form of a cell, is present. When cells are not being created for a given source-destination pair, the

same network resources are used to transmit cells for other source-destination pairs. Cells are packets of fixed length and comprise both flow control (i.e., cell header) and payload information.

Any or each of the nodes in a telecommunications network may comprise a cell memory or buffer which is available to a number of traffic flows. These buffers may exist at various levels, for example, at the port level, the card level (where a single card supports multiple ports), the switch level, the class of service level, etc. As used hereafter, the term "per flow" is meant to include any or all situations where a single buffer resource is shared by a number of traffic flows, regardless of the level on which the sharing may occur.

The basic operation for per flow buffer control according to the present invention is described with reference to Figures 1a-1c. As shown in Figure 1a, an initialization procedure begins at step 10. At step 12, a shared buffer is initialized and a buffer count and flow cell counts are reset to zero. The use of these counts is described below.

Figure 1b illustrates the operation of dynamic threshold scaling for a preferred embodiment. As a new cell arrives at step 14, the corresponding traffic flow is determined from the cell header information at step 16. The default threshold and buffer count for that flow are then retrieved from memory at step 18. The default threshold can be thought of as representing the maximum amount of buffer resources that a flow may use if no other flows are currently using the buffer. In other words, the default threshold represents the maximum number of cells a given flow may store in the buffer under the "ideal" condition where no other traffic flows or resources are using the buffer. These thresholds may be determined based on factors such as total available buffer size, customer requirements, traffic type, etc. The flow cell count represents the number of cells corresponding to the particular traffic flow of interest which are already stored in the buffer. At step 20, the total buffer utilization is determined. That is, the total number of cells from all traffic flows which are stored in the buffer is determined. Using the global buffer utilization from step 20 and the flow identification from step 16 as indices, a scaling factor for the flow threshold is retrieved from a lookup table stored in memory at step 22. The scaling factor is used to calculate the flow dynamic threshold in step 24 according to the following formula:

-6-

$$\text{Dynamic Thresh}_i = \text{Thresh}_i * \text{SF}_i$$

where:

Dynamic Thresh_i = the dynamic threshold for the i^{th} traffic flow;

Thresh_i = the default threshold for the i^{th} traffic flow; and

SF_i = the scaling factor for the i^{th} traffic flow according to the global buffer utilization.

At step 26 a comparison is made to determine if the number of cells corresponding to the traffic flow of interest already stored in the buffer exceeds the dynamic threshold for that flow. If so, the process moves to step 28 and the new cell is dropped. Otherwise, the process moves to step 30 where the new cell is admitted and the buffer count for the flow of interest and the global buffer count are incremented.

As shown in Figure 1c, when a cell departs the buffer, the corresponding flow cell count and the global buffer counts are decremented. In this way, current buffer counts are maintained at both the flow and global levels. The process 100 of Figure 1 is further described in detail with reference to Figures 2-6, below.

Referring to Figure 2, a graph depicting the common usage of a single buffer resource by a number of traffic flows is shown. The horizontal axis of the graph of Figure 2 shows the traffic flows which are sharing the buffer. Although Figure 2 shows only five flows sharing the single buffer, those skilled in the art will appreciate that this is for purposes of clarity and simplicity only and that the buffer management methods of the present invention are equally applicable to situations where any number of traffic flows share a single common buffer. The vertical axis of the graph shown in Figure 2 is a count of the number of cells stored in the buffer by each flow. For the example shown in Figure 1, traffic flow 1 has 500 cells stored, traffic flow 2 has 1250 cells stored, traffic flow 3 has 750 cells stored, traffic flow 4 has 650 cells stored and traffic flow 5 has 1000 cells stored. Thus, for the example shown in Figure 2, a total of 4150 cells are stored in the shared buffer.

Figure 3 further illustrates the example begun in Figure 2. Each flow has the same number of cells stored in the shared buffer as in Figure 2. In Figure 3, however, a number of default thresholds are shown. Each default threshold (Thresh_1 through Thresh_5) corresponds to a respective one of the traffic flows 1 through 5. For the example shown in Figure 3, Thresh_1 is set at 3000 cells,

Thresh₂ is set at 9000 cells, Thresh₃ is set at 4000 cells, Thresh₄ is set at 5000 cells and Thresh₅ is set at 7000 cells. As indicated above, the default thresholds represent the maximum amount of buffer resources that each particular flow may use if no other connections are currently using the buffer. The thresholds have been determined based on factors such as total available buffer size, customer requirements, traffic type, etc.

Referring now to Figure 4, the graph of buffer utilization for the shared buffer is shown at a later point in time than was depicted in Figures 2 and 3. In the situation depicted in Figure 4, flows 2 and 4 have added a number of cells to the common buffer. Traffic flow 2 now has 5,500 cells stored in the buffer and traffic flow 4 has 1250 cells stored in the buffer. Flows 1, 3 and 5 have neither added nor removed cells from the buffer. Thus, a total of 9000 cells are stored in the buffer for the instant of time shown in Figure 4. If the common buffer is capable of storing a maximum of 10,000 cells total, for the example depicted in Figure 4 the buffer is at 90% capacity.

Suppose now a new cell arrives. In accordance with the methods of the present invention, the new cell's corresponding traffic flow information is determined. For this example, suppose the new cell is associated with flow 2. The default threshold and buffer count for flow 2 are retrieved from memory. As shown in Figure 4, Thresh₂ is 9000 cells and the global buffer count is 9000 cells (i.e., 90% of capacity). Using the global buffer utilization, the appropriate scaling factor for the flow 2 threshold is retrieved from a lookup table stored in memory. For this example, suppose the flow 2 threshold is to be scaled back to one-half of its default value when global buffer utilization reaches 90% of capacity (the very situation depicted in Figure 4). Thus, the Dynamic Thresh₂ based on the current buffer utilization is $1/2 * 9000 = 4500$ cells. This is graphically illustrated in Figure 5.

The decision on whether to admit the new cell for flow 2 is now based on the dynamically scaled threshold for flow 2. As shown in Figure 5, flow 2 is already storing 5500 cells in the common buffer. This exceeds the dynamically scaled threshold for this flow (which is 4500 cells for the depicted buffer utilization condition). As a result, the new cell is dropped.

Suppose now a new cell corresponding to traffic flow 1 arrives. The buffer utilization has not changed from the situation depicted in Figures 4 and 5.

-8-

Flow 1 still stores 500 cells, flow 2 is storing 5500 cells, flow 3 is storing 750 cells, flow 4 is storing 1250 cells and flow 5 is storing 1000 cells. Thus, buffer utilization remains at 90% of capacity. Also, recall that the default threshold for flow 1 is 3000 cells, as shown in Figure 3.

As the new cell for flow 1 arrives, the default threshold (3000 cells) and current buffer count (500 cells) for flow 1 are retrieved from memory. Using the global buffer utilization (90%), the appropriate scaling factor for the flow 1 threshold is retrieved from the lookup table stored in memory. For this example, suppose that like flow 2, the flow 1 threshold is to be scaled back to one-half of its default value when global buffer utilization reaches 90% of capacity. Thus, the Dynamic Thresh₁ based on the current buffer utilization is $1/2 * 3000 = 1500$ cells. This is graphically illustrated in Figure 6.

The decision on whether to admit the new cell for flow 1 is now based on the dynamically scaled threshold for flow 1. As shown in Figure 6, flow 1 is only storing 500 cells in the common buffer. This is less than the number of cells permitted by the dynamically scaled threshold for this flow. As a result, the new cell is admitted to the buffer and the global buffer count and flow 1 buffer count are incremented.

The above examples illustrate how dynamic scaling penalizes only those flows which are using significant portions of their allocated buffer space. Flow 2 was using a significant portion of its allocated capacity (5500 out of 9,000). As the buffer reached the 90% full level, dynamic scaling was employed and flow 2 was not permitted to store any more cells. Under these conditions, flow 2 would not be allowed to store any more cells until global buffer utilization had declined. On the other hand, given the same buffer utilization (90% of capacity), flow 1, which was storing only 500 cells (1/6 of its configured maximum), was permitted to store another cell. Note also that although flow 2 was using a significant amount of buffer space, dynamic scaling only affected newly arriving cells. That is, although flow 2 was already storing more cells (5500) than would otherwise be permitted according to the dynamically scaled threshold (4500), no previously stored cells were discarded.

It will be appreciated that different scaling tables can be provided for different traffic flows. For instance, scaling Class 1 might be used for UBR traffic, Class 2 might be for ABR traffic and Classes 3 and 4 used for more

sensitive traffic such as VBR and CBR. Table 1 shows some exemplary settings, although it will be appreciated that other scaling factors could be used. The per flow thresholds for ABR and UBR traffic are likely to be set aggressively high as these classes can tolerate scaling back early. Other traffic types (such as CBR and VBR) would generally have smaller per flow thresholds but would be more sensitive to scaling back.

Table 1

Buffer Utilization	Scaling Factor Class 1	Scaling Factor Class 2	Scaling Factor Class 3	Scaling Factor Class 4
< 90%	1	1	1	1
90%	1/2	1	1	1
91%	1/4	1	1	1
92%	1/8	1/2	1	1
93%	1/16	1/4	1	1
94%	1/32	1/8	1	1
95%	1/64	1/16	1	1
96%	1/256	1/32	1/2	1
97%	1/1024	1/64	1/2	1
98%	1/4096	1/256	1/4	1
99%	1/16384	1/16384	1/4	1

The 1% increments, the starting value of 90%, and the scaling fractions are all examples only. The contents of the table are, in general, configurable. For example, to provide a safety margin for CBR and VBR queues, it may be desirable to move the scale table lower, that is, replacing the 9X% with 8X% or 7X%. Also, the scaling factors can be made user selectable based on network conditions and/or requirements.

Limiting the scaling factors to binary fractions can drastically simplify the implementation. In such an embodiment, the thresholds are preferably stored in a format having a 4-bit mantissa (M) and a common 4-bit exponent (E). The linear threshold (T) is calculated as $T=M \times 2^E$. Thus, the scaling can be easily achieved by adjusting the exponent such that $T=M \times 2^{(E-A)}$ where A is obtained from Table 1

-10-

(in other words, Table 1 would actually store the adjustment $A = \{0, 1, 2, 3, \dots\}$ rather than the fraction $\{1, 1/2, 1/4, 1/8, \dots\}$).

It will be appreciated that the dynamic threshold scaling method has several performance advantages over existing techniques. For example, the method is scalable to a large number of traffic flows and for a large number of per flow queues. Sensitive traffic flows can be isolated from "memory hogs". The method further ensures "fair" allocation of resources between flows in the same scale and class. Note that "fair" does not necessarily mean equal (at 90% buffer utilization, flow 2 was permitted to store 4500 cells while flow 1 was only allowed 1500), rather, resource allocation may be determined by individual customer needs. Dynamic scaling further allows preferential treatments of groups of traffic flows via the selection of scaling classes. Global resource overflows are avoided and, hence, the performance degradation that accompanies these events is avoided.

Storing the thresholds associated with the flow in the form of separate mantissas with a shared exponent drastically reduces the memory which would otherwise be required to store these thresholds. A conventional approach would require 20 bits per threshold per flow. The preferred method, however, requires just 20 bits to store all the thresholds (assuming a 4-bit representation). This makes a significant difference when the number of flows is large. Furthermore, it reduces the processing bandwidth because the buffer count comparisons share a common exponent test and a simple 4-bit mantissa comparison.

In addition to maximum threshold discard control as described above, a number of other options can be supported using the methods of the present invention. For example, decisions on whether to discard cells which have their cell loss priority (CLP) bits set or decisions on whether to use EFCI congestion control can be made according to dynamically scaled thresholds. Such implementations are discussed below.

Typically, each node in a network maintains information regarding each traffic flow (e.g., VP and/or VC) it supports. To implement per flow dynamic scaling management options, additional information would be maintained by these nodes. Then, for each cell, the cell header is used to generate a flow indicator that indexes a lookup table that contains information regarding the traffic flow of interest. The lookup table may be stored in a memory associated with the network

node of interest and would store a number of thresholds which could be dynamically scaled.

As an example, consider the case where a CLP threshold is to be scaled. In such an embodiment, cells which have their CLP bits set (CLP = 1) will be discarded when the buffer count for the traffic flow of interest exceeds the CLP threshold. The CLP thresholds can be dynamically scaled according to the procedure described above.

As another example, consider the case of frame discard. In many applications, data is sent in frames. In such cases, once one cell is lost, the rest of the cells in the frame are not useful. "Goodput" can therefore be improved by discarding the remaining cells in the frame. A single bit per flow (frame discard enable bit) could be used to enable this feature.

In this case, logic associated with the node containing the common buffer would keep track of end-of-frame (EOF) indicators in arriving cells. In this way, frames could be distinguished. Various state information determined from the EOF indicators and a dynamically scaled early packet discard threshold could then be used to trigger frame discarding.

Per flow buffer management can also be used to set the EFCI bit in cell headers to allow for the use of other congestion management processes. The EFCI threshold is checked as cells are serviced. If the buffer count for the traffic flow of interest is greater than the EFCI threshold for that flow, the EFCI bit in the cell header set. Again, the EFCI threshold can be dynamically scaled according to the above described process.

Thus, an efficient method for managing a common communications buffer resource shared by a large number of traffic flows (e.g., processes or connections) has been described. In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will, however, be clear that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. For example, those skilled in the art will appreciate that the common buffers described in the specification may exist at a variety of levels within a network switch node. This includes the port level, card level, switch level, etc. Also, exemplary thresholds of interest, such as maximum cell discard thresholds, CLP thresholds, EPD thresholds and EFCI thresholds have been

-12-

discussed. Although discussed separately, those skilled in the art will recognize that the buffer count checks for each of these thresholds may be performed simultaneously or in various groupings according to user and network requirements. Further, these are only examples of the types of thresholds which might be dynamically scaled. Those skilled in the art will recognize that a number of other thresholds may be dynamically configured to achieve desired traffic management in a data communications network. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A method of managing a common buffer resource shared by a plurality of processes including a first process, the method comprising the steps of:

establishing a first buffer utilization threshold for said first process;
monitoring the usage of said common buffer by said plurality of processes;
and

dynamically adjusting said first buffer utilization threshold according to
said usage.

2. The method of claim 1 wherein said plurality of processes is a
plurality of traffic flows and said first process is a first traffic flow.

3. The method of claim 1 wherein said first buffer utilization threshold
is represented in a format having a mantissa and an exponent.

4. The method of claim 1 wherein said first buffer utilization threshold
is a cell maximum threshold.

5. The method of claim 1 wherein said first buffer utilization threshold
is a cell loss priority (CLP) threshold.

6. The method of claim 1 wherein said first buffer utilization threshold
is an early packet discard (EPD) threshold.

7. The method of claim 1 wherein said common buffer has a
maximum capacity and wherein the step of dynamically adjusting comprises:

determining said usage of said common buffer by said plurality of
processes;

determining a first scaling factor according to said usage of said common
buffer; and

scaling said first buffer utilization threshold by said first scaling factor.

-14-

8. The method of claim 3 wherein the step of dynamically adjusting comprises performing a subtraction operation on said exponent.

9. The method of claim 3 wherein a plurality of thresholds, each of said plurality of thresholds corresponding to a respective one of said plurality of processes, share a common exponent and wherein said plurality of thresholds are scaled simultaneously using a subtraction operation on said common exponent.

10. A buffer management process for a cell switching communications network having a first node, said first node having a common buffer being shared by a plurality of network connections including a first connection, the process comprising the steps of:

receiving a first cell at said first node, said first cell being associated with said first connection;

determining a buffer count for said common buffer, said buffer count representing a current utilization of said common buffer by said plurality of connections;

establishing a first connection threshold for said first connection according to said buffer count; and

determining whether said first cell will be accommodated in said common buffer using said first connection threshold.

11. The buffer management process of claim 10 wherein the step of establishing the first connection threshold comprises:

establishing an initial connection threshold; and

dynamically adjusting said initial connection threshold according to said buffer count, said dynamic adjusting producing a first scaled threshold.

12. The buffer management process of claim 11 wherein the step of determining whether said first cell will be accommodated comprises:

establishing a connection cell count, the connection cell count indicating the number of cells associated with the first connection stored in said common buffer;

comparing said connection cell count to said first scaled threshold,

wherein if said connection cell count exceeds said first scaled threshold, said first cell is not admitted to said common buffer.

13. The buffer management process of claim 12 wherein said first connection threshold is coded in a format having a mantissa and an exponent and said step of dynamically adjusting comprises a subtraction operation.

14. A buffer management system for congestion prevention in a cell switching communications network comprising a plurality of logical connections, the buffer management system comprising:

a first node receiving network traffic transmitted over said plurality of logical connections, said first node having a common buffer and further having a buffer control device, said buffer control device monitoring the usage of said common buffer by said network traffic and dynamically scaling a buffer utilization threshold according to said usage, said buffer utilization threshold corresponding to a first of said plurality of logical connections.

15. A buffer management system as in claim 14 wherein said buffer control device further comprising:

a lookup table stored in a memory, said lookup table comprising buffer threshold scaling factors.

16. The method of claim 14 wherein said buffer utilization threshold is represented in a format having a mantissa and an exponent.

17. A buffer management system as in claim 14 wherein said buffer utilization threshold is a cell loss priority (CLP) threshold.

18. A buffer management system as in claim 14 wherein said buffer utilization threshold is early packet discard (EPD) threshold.

19. A buffer management system as in claim 14 wherein said buffer utilization threshold is a cell maximum threshold.

20. A buffer management system as in claim 15 wherein said buffer control device further comprising circuitry for comparing, said circuitry for comparing receiving a first signal indicating a buffer count, the buffer count representing the utilization of said common buffer by said first connection, said circuitry for comparing further receiving a second signal representing said scaled

-16-

buffer utilization threshold after dynamic scaling, said circuitry for comparing generating a third signal, said third signal indicating whether said first connection has exceeded an associated allowable usage of said common buffer..

21. A cell exchange node for a cell switching communications network, the node comprising:

a lookup table stored in a memory, said lookup table comprising buffer threshold scaling factors.

22. A cell exchange node as in claim 21 wherein said communications network comprising a plurality of logical connections and wherein said memory further comprising a plurality of connection thresholds, said plurality of connection thresholds each representing a maximum number of cells to be stored in a buffer associated with said cell exchange node for each respective logical connection.

23. A cell exchange node as in claim 22 wherein said plurality of connection thresholds are coded in a format having a mantissa and an exponent.

24. A cell exchange node as in claim 23 wherein said plurality of connection thresholds share a common exponent.

1/8

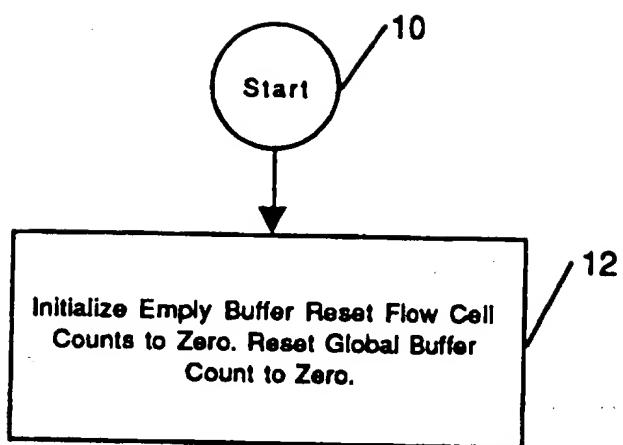
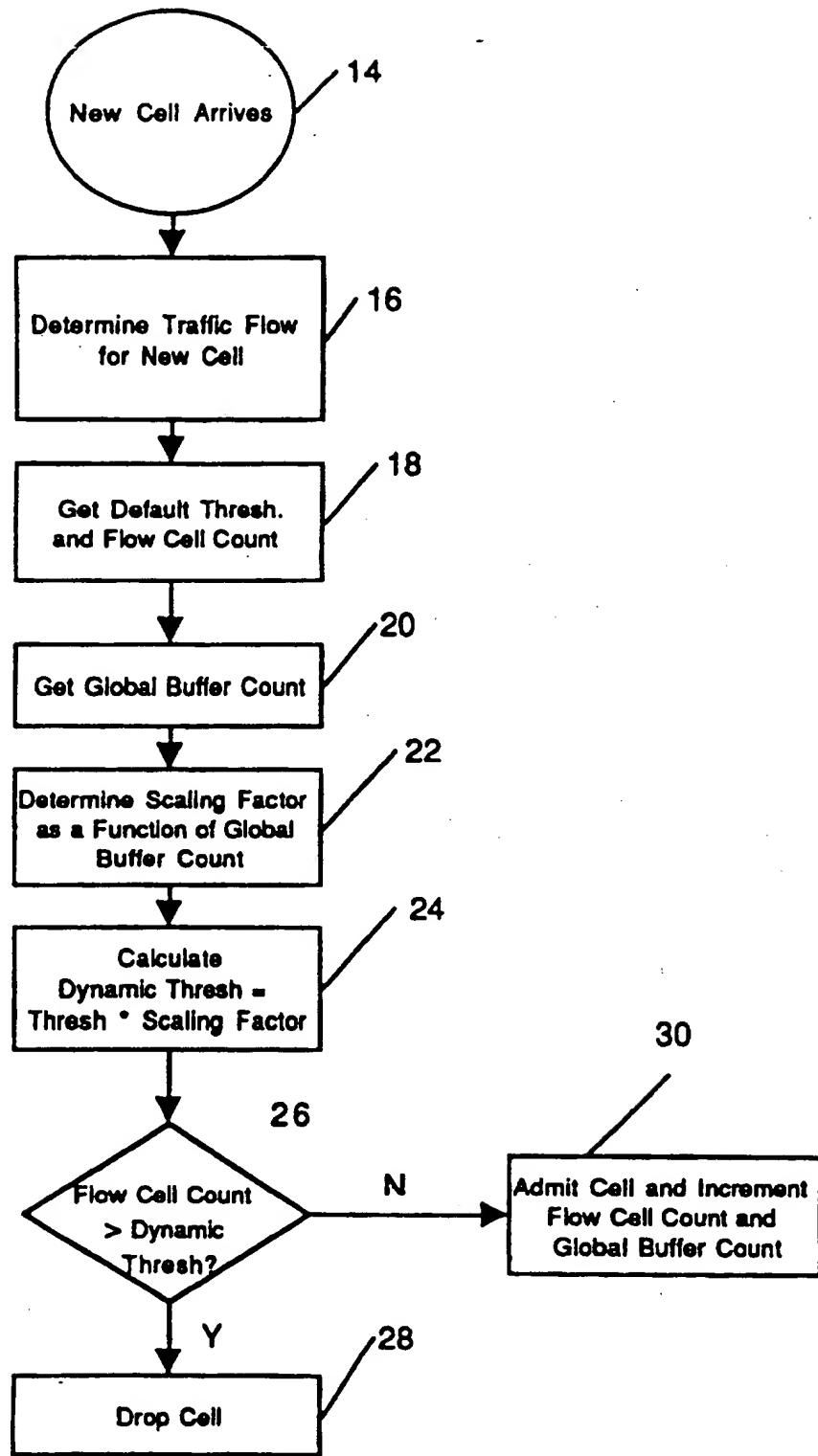


FIGURE 1a

**FIGURE 1b**

3/8

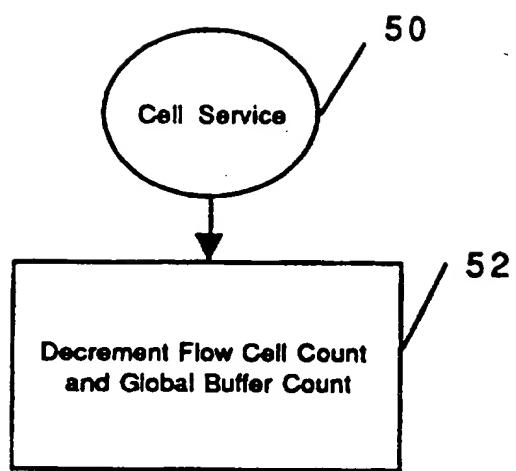


FIGURE 1c

4/8

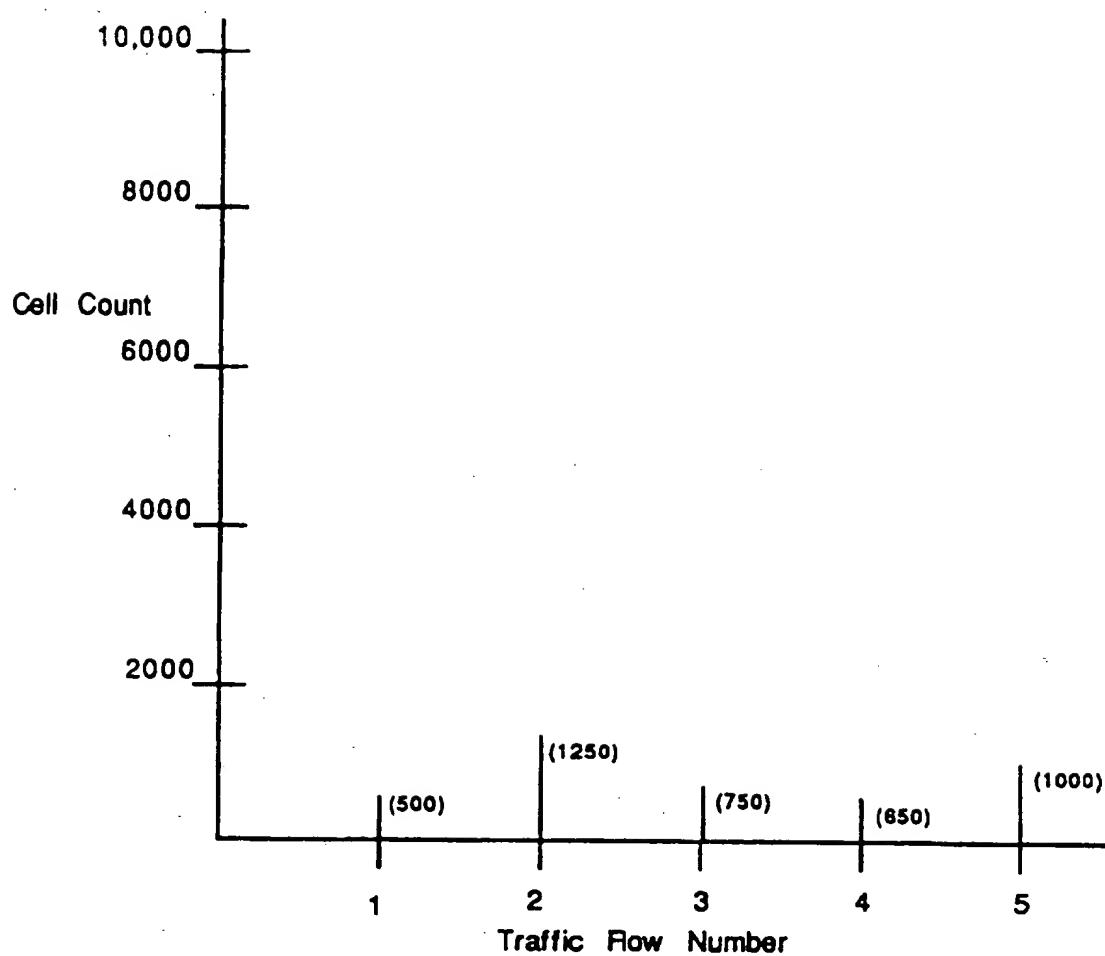


FIGURE 2

5/8

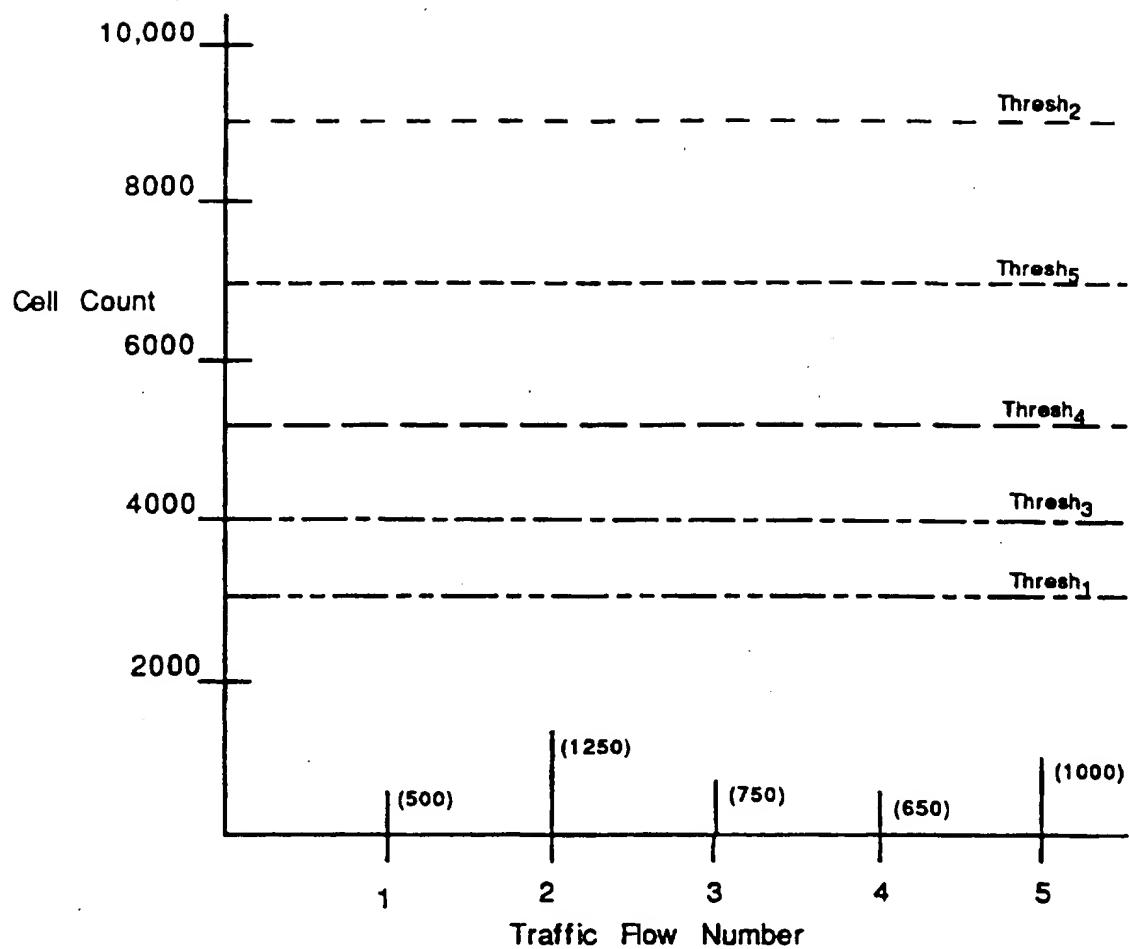


FIGURE 3

6/8

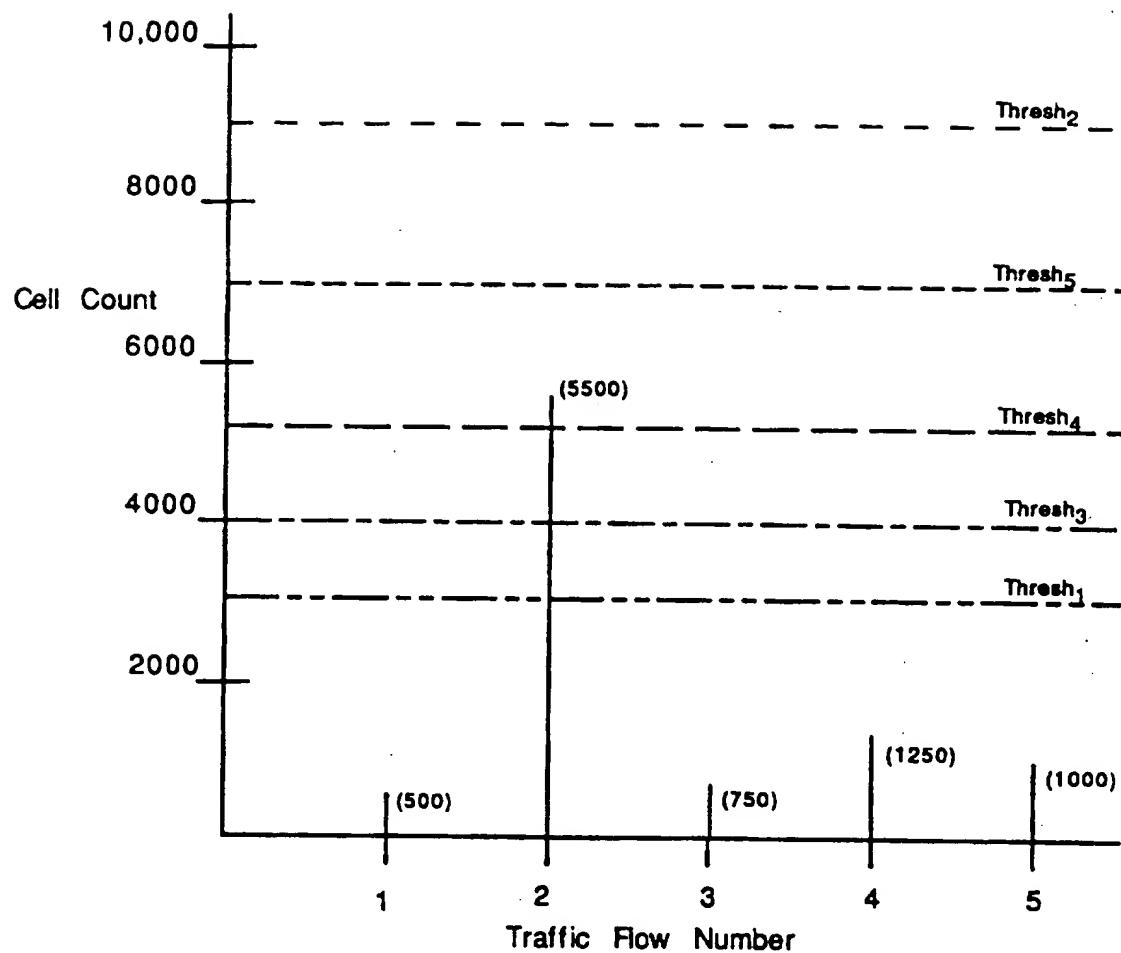


FIGURE 4

7/8

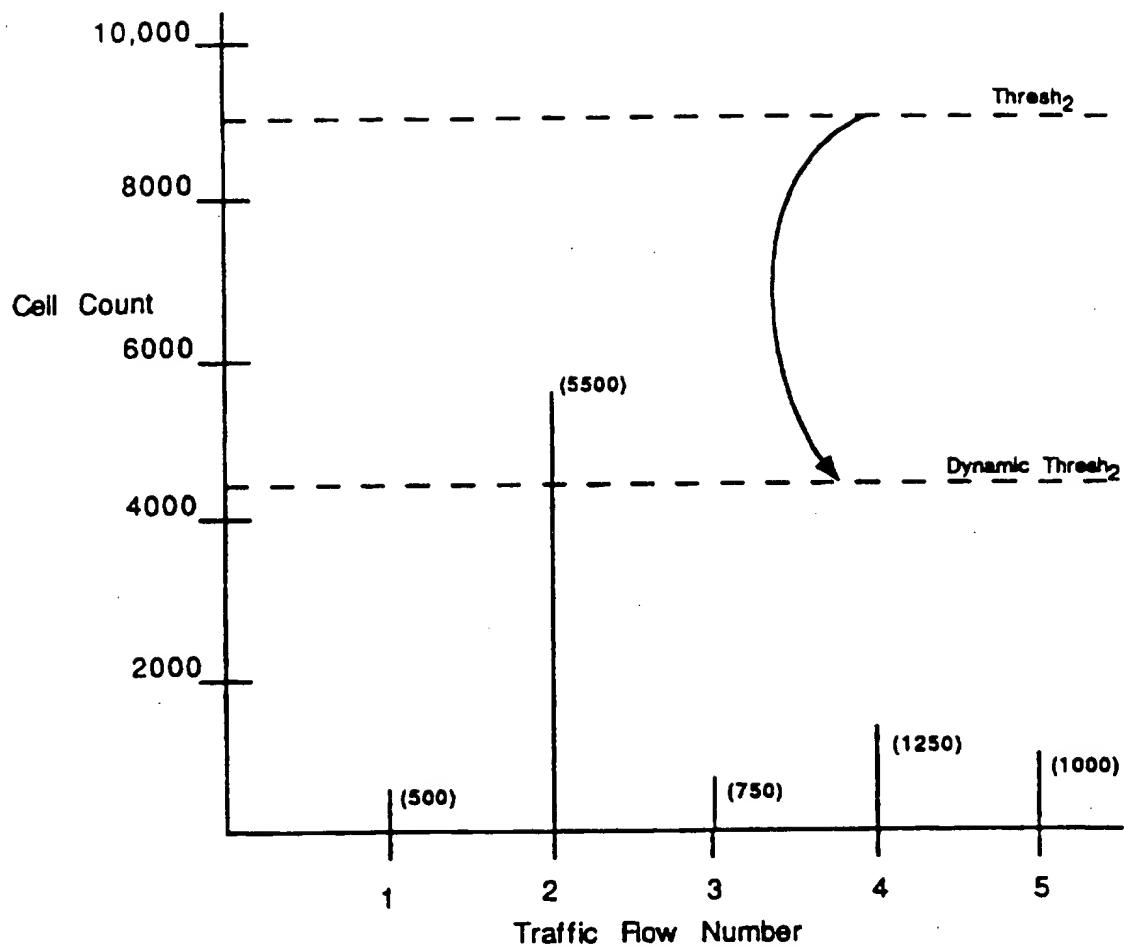


FIGURE 5

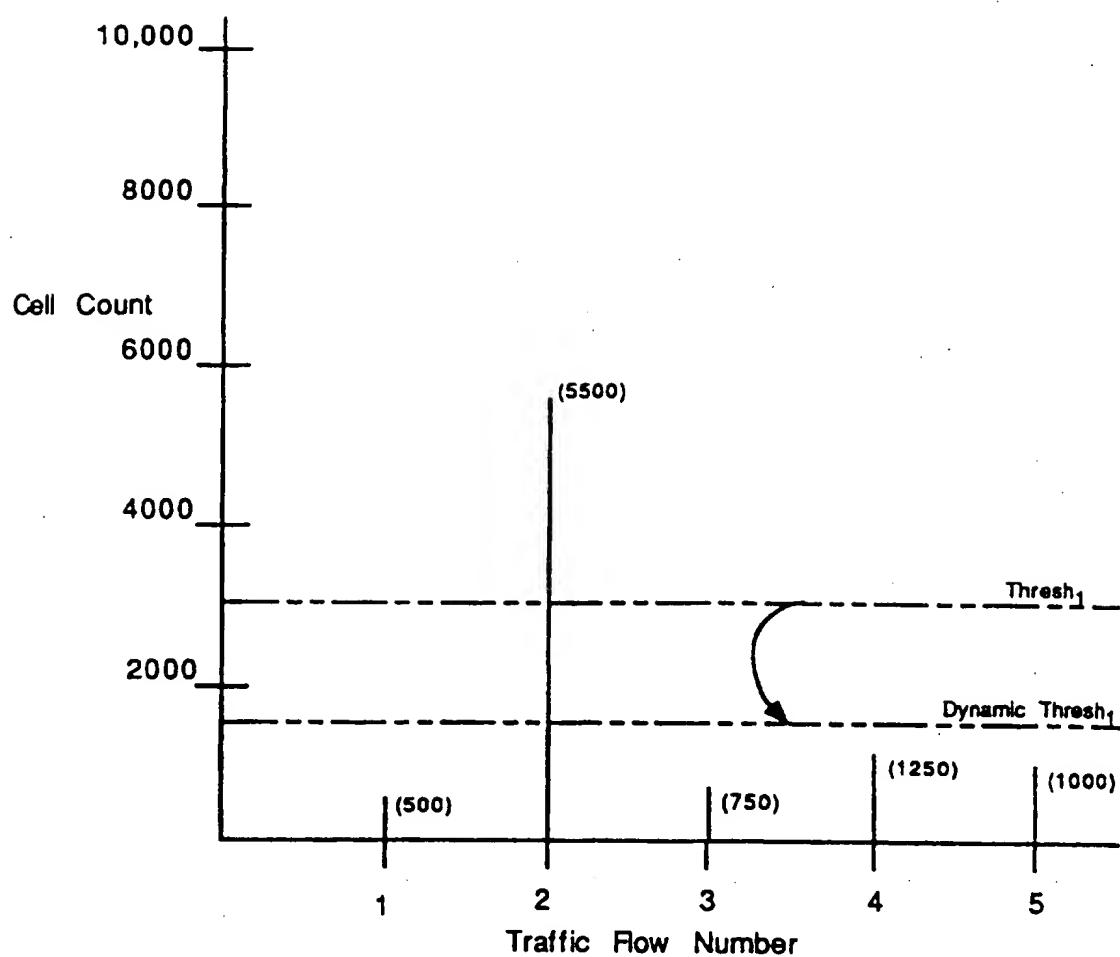


FIGURE 6

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/07839

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 6 H04Q11/04 H04L12/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 IPC 6 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 706 298 A (AT & T CORP) 10 April 1996 see page 2, line 36 - page 3, line 3 see page 3, line 51 - page 4, line 18 see page 4, line 37 - line 50	1-4,7, 10-16, 19-23
Y	see page 5, line 16 - line 31 see page 6, line 53 - line 55 ---	5,6,17, 18
A	US 5 434 848 A (CHIMENTO JR PHILIP F ET AL) 18 July 1995 see column 6, line 1 - line 52 ---	8,9,24
Y	-/-	5,17

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- 'A' document defining the general state of the art which is not considered to be of particular relevance
- 'E' earlier document but published on or after the international filing date
- 'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- 'O' document referring to an oral disclosure, use, exhibition or other means
- 'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

'&' document member of the same patent family

2

Date of the actual completion of the international search

9 September 1997

Date of mailing of the international search report

18.09.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl.
 Fax (+ 31-70) 340-3016

Authorized officer

Gregori, S

INTERNATIONAL SEARCH REPORT

Intern: Application No

PCT/US 97/07839

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KENJI KAWAHARA ET AL: "PERFORMANCE EVALUATION OF SELECTIVE CELL DISCARD SCHEMES IN ATM NETWORKS" PROCEEDINGS OF IEEE INFOCOM 1996. CONFERENCE ON COMPUTER COMMUNICATIONS, FIFTEENTH ANNUAL JOINT CONFERENCE OF THE IEEE COMPUTER AND COMMUNICATIONS SOCIETIES. NETWORKING THE NEXT GENERATION SAN FRANCISCO, MAR. 24 - 28, 1996, vol. 3, 24 March 1996, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 1054-1061, XP000622238 see page 1054, column 1, line 47 - column 2, line 10 -----	6,18

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat'l Application No

PCT/US 97/07839

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0706298 A	10-04-96	US 5541912 A CA 2156654 A JP 8251233 A	30-07-96 05-04-96 27-09-96
US 5434848 A	18-07-95	NONE	

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.